ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN — POLONIA
VOL. XXIII, 27 SECTIO H 1989

Zakład Nauk Ekonomicznych
Wydziału Ekonomicznego Filii UMCS w Rzeszowie

Adam GÓRAL

Badanie stacjonarności jednowymiarowych procesów losowych

Стационарные исследование однамерных случайных процессов

The Study of Stationarity of One-dimensional Stochastic Processes

Modele ARIMA (p,d,q) są obecnie szeroko stosowane do opisu dynamiki prognozowania procesów losowych, pomimo że metody identyfikacji [1] tych modeli są niejednoznaczne i często wysoce wątpliwe. Podstawę identyfikacji wspomnianych modeli stanowią uwagi zawarte w pracy G. E. P. Boxa i G. M. Jenkinsa [6, s. 174—187], zgodnie z którymi ocena parametrów p, d, q może być dokonana w oparciu o analizę wykresów ocen wartości funkcji autokorelacji i autokorelacji cząstkowej. W niniejszej pracy uwagę skoncentrowano na problemie doboru rzędu różnicowania d, czyli na badaniu stacjonarności w szerszym sensie procesu reprezentowanego przez określony szereg czasowy.

Praca składa się z sześciu części i zakończenia. Po uwagach wstępnych, w drugiej części pracy przedstawiono uwagi dotyczące estymacji podstawowych charakterystyk procesów losowych. W części trzeciej omówiono istotę zaproponowanej przez Boxa i Jenkinsa metody badania stacjonarności. Część czwartą poświęcono wykorzystaniu medianowego testu serii do badania stacjonarności jednowymiarowych ciągów losowych. W części piątej dokonano oceny efektywności metod badania stacjonarności, które są stosunkowo często wykorzystywane w praktyce, czyli metody Boxa i Jenkinsa oraz medianowego testu serii. Ocenę tę przeprowadzono w oparciu o wyniki badań symulacyjnych. W szóstej części pracy zaproponowano metodę badania stacjonarności, która — jak sądzi autor — umożliwia bardziej wiarygodną ocenę stacjonarności, niż metody dotychczas stosowane.

[1] Identyfikacja modeli ARIMA (p, d, q) polega na ocenie parametrów p, d i q.

ESTYMACJA CHARAKTERYSTYK SŁABO STACJONARNYCH PROCESÓW LOSOWYCH

Pod pojęciem słabo stacjonarnego procesu losowego rozumiemy proces stacjonarny ze względu na wartość oczekiwaną i funkcję kowariancji, czyli:

$$\bigwedge_{t=0,\pm 1,\pm 2,\ldots} E(X(t))=\mu \quad (1)$$

$$\bigwedge_{t_1,t_2,t_3,t_4=0,\pm 1,\pm 2,\ldots} |t_1-t_2|=|t_3-t_4| \Leftrightarrow cov(X(t_1)X(t_2)) \Leftrightarrow cov(X(t_3),X(t_4)). \quad (2)$$

W przypadku stacjonarności procesu ze względu na funkcję kowariancji mamy możliwość[2] wprowadzenia pojęcia funkcji autokowariancji o postaci:

$$\psi(\tau)=cov(X(t_i),X(t_j)), \quad (3)$$

gdzie: $\tau=t_i-t_j (\tau=0,\pm 1,\pm 2,\ldots)$

Z własności (2) wynika parzystość funkcji autokowariancji, czyli $\psi(\tau)=\psi(-\tau)$.

Przebieg procesów losowych charakteryzowany jest często na podstawie ich funkcji korelacji, czyli

$$\varrho(t_i,t_j)=cov(X(t_i),,X(t_j))/(s(t_i)\cdot s(t_j)), \quad (4)$$

gdzie:

$$s(t_i)=\sqrt{cov(X(t_i),X(t_i))}$$

Proces losowy jest stacjonarny ze względu na funkcję korelacji, gdy funkcję $\varrho(t_i,t_j)$ można zastąpić funkcją jednej zmiennej $\tau=t_i-t_j(t_j,t_j=0, \pm 1, \pm 2,\ldots)$. Funkcję $\varrho(\tau)$ nazywamy funkcją autorelacji. Przedstawione uwagi wskazują, że stacjonarność ze względu na funkcję korelacji nie implikuje stacjonarności ze względu na kowariancję[3].

Omówione wyżej charakterystyki szacowane są w przypadku założenia o słabej stacjonarności i ergodyczności danego procesu losowego na podstawie następujących wzorów:

$$\bar{x}=n^{-1}\sum_{t=1}^{n} x_t, \quad (5)$$

---

[2] W pracy świadomie użyto słów „mamy możliwość". W wielu pracach z zakresu analizy szeregów czasowych pojęcie funkcji autokorelacji jest bowiem używane dla procesów niestacjonarnych ze względu na funkcję kowariancji. Tego typu podejście jest — według autora niniejszej pracy — niezgodne z teorią procesów losowych.

[3] Uwaga ta wynika z faktu, że można wyobrazić sobie taki proces losowy, dla którego $\varrho(t_i, t_j)$ nie zależy od t_i, t_j, lecz od $\tau=t_i-t_j$, pomimo, iż proces ten nie jest stacjonarny ze względu na funkcję kowariancji.

$$c(\tau)=n^{-1}\sum_{t=1}^{n-\tau}(x_t-\bar{x})(x_{t+1}-\bar{x}), \qquad \tau=0,1,...,m \tag{6}$$

$$r(\tau)=c(\tau)/c(0), \qquad \tau=0,1,...,m \tag{7}$$

gdzie: $\{x_1,x_2,...,x_n\}$ oznacza pojedynczą realizację analizowanego procesu losowego,

m jest punktem odcięcia funkcji autokowariancji i autokorelacji.

Średni błąd szacunku parametru $\varrho(\tau)$ oceniany jest w oparciu o formułę Bartletta [6, s. 179]. Formuła ta przedstawiona jest w następujący sposób:

$$s(r(k))=n^{-1/2}(1+2(r^2(1)+r^2(2)+...+r^2(q)))^{1/2}, \qquad k>q \tag{8}$$

Wartości funkcji autokoleracji uznawane są jako statystycznie nieistotne, gdy

$$r(k)\in\langle -2\cdot s(r(k)),2\cdot s(r(k))\rangle. \tag{9}$$

OCENA RZĘDU RÓŻNICOWANIA W METODZIE BOXA-JENKINSA

Załóżmy, że do szeregu czasowego $x_1,x_2,...,x_n$ stanowiącego realizację procesu losowego $\{X(t);$
$t=0,\pm 1,\pm 2,...\}$ dopasowujemy model ARIMA, p, d, q) o postaci:

$$F(B)\nabla^d x_t=Q(B)a_t, \tag{10}$$

gdzie:

$F(B)=1-\alpha_{p1}B-\alpha_{p2}B^2-...-\alpha_{pp}B^p,$

$Q(B)=1-h_{q1}B-h_{q2}B^2-...-h_{qq}B^q,$

$\nabla^d x_t=(1-B)^d x_t,$

$Bx_t=x_{t-1}, \quad Ba_{t-1},$

$B^p x_t=x_{t-p} \quad B^q a_t=a_{t-q},$

$\{a_t\}$ oznacza proces czysto losowy o wariancji δ^2_a.

Identyfikacja modeli (10) polega na ocenie parametrów p, d, q oznaczających odpowiednio: rząd składowej autoregresyjnej (AR), rząd różnicowania doprowadzającego proces $\{X_t\}$ do słabej stacjonarności i rząd składowej występującej w postaci średniej ruchomej (MA).

G. E. P. Box i G. M. Jenkins proponują, by rząd różnicowania oznaczał taką najmniejszą liczbę z ciągu 0, 1, 2,..., dla której szereg czasowy $w_t=\nabla^d x_t$można uznać za realizację słabo stacjonarnego procesu losowego. Wspomniani autorzy sugerują, że decyzja odnośnie stacjonarności (bądź niestacjonarności) danego procesu losowego może być podejmowana w oparciu o obserwację „zachowania" ocen wartości funkcji autokorelacji danego procesu. Powyższa sugestia wynika z następującego stwier-

dzenia [6, s. 175]: „w przypadku modelu stacjonarnego, w którym żaden z pierwiastków nie leży blisko okręgu jednostkowego funkcja autokorelacji szybko zanika dla dużych i średnich k". Stąd wyciągany jest wniosek, że w sytuacji, gdy funkcja autokorelacji nie wykazuje tendencji do szybkiego zaniku analizowany szereg czasowy można rozważać jako realizację procesu niestacjonarnego. Box i Jenkins podejmują próbę formalnego wyjaśnienia przedstawionych wyżej spostrzeżeń. Podstawę tego wyjaśnienia stanowi fakt, że funkcja autokorelacji słabo stacjonarnego procesu losowego ARMA (p, q) spełnia równanie różnicowe o postaci:

$$\varphi(B)\varrho(k)=0 \qquad k>q, \qquad (11)$$

gdzie

$$\varphi(B)=\prod_{i=1}^{p}(1-G_iB).$$

Przy założeniu, że wszystkie pierwiastki równania różnicowego (11) są jednokrotne, jego rozwiązanie może być przedstawione w następujący sposób:

$$\varrho(k)=A_1G^k_1+A_2G^k_2+\ldots+A_pG^k_p \qquad k>q-p, \qquad (12)$$

Z równania (12) wynika, że w przypadku, gdy przynajmniej jeden pierwiastek równania charakterystycznego leży w pobliżu okręgu jednostkowego, funkcja autokorelacji opada wolno ([6, s. 176]). Uważam, że równanie (12) nie może być wykorzystywane do wyciągania jakichkolwiek wniosków odnośnie przebiegu teoretycznej funkcji autokorelacji procesu niestacjonarnego, gdyż funkcja taka nie istnieje. Można jedynie podjąć próbę analizy kształtowania się wartości estymatora funkcji autokorelacji procesu stacjonarnego w sytuacji, gdy dysponujemy realizacją procesu niestacjonarnego. Wszelkie uogólnienia mogą być dokonywane jedynie na podstawie wyników badań symulacyjnych.

MEDIANOWY TEST SERII W BADANIU STACJONARNOŚCI JEDNOWYMIAROWYCH CIĄGÓW LOSOWYCH.

Obok metody Boxa i Jenkinsa podstawę badania stacjonarności procesów losowych stanowi również medianowy test serii omówiony m. in. w [4] i [7]. Zgodnie z propozycją J. S. Bendata i A. G. Piersola badanie stacjonarności sprowadzane jest do oceny stacjonarności ze względu na wartość średnią i wariancję. Zastosowanie wspomnianego wyżej testu do weryfikacji hipotezy o stacjonarności ze względu na średnią wymaga podziału n elementowego szeregu czasowego (realizacja procesu) na k rozłącznych s elementowych podciągów $\{x_1, x_2,\ldots, x_s\}$, $\{x_{s+1}, x_{s+2},\ldots, x_{2s}\},\ldots$, $\{x_{(k-1)\cdot s+1},\ldots,x_{k\cdot s}\}$. Dla każdego podciągu wyznaczane są wartości

$$\bar{x}_i = s^{-1} \sum_{j=1}^{s} x_{(i-1)\cdot s+j} \qquad (i=1, 2,...,k), \qquad (13)$$

Przy założeniu prawdziwości H_0 o stacjonarności ze względu na średnią, wartości $\bar{x}_i$; $i=1, 2,...,k$ stanowią realizacje zmiennej losowej o wartości oczekiwanej μ. Widać więc, że ciąg wartości $\bar{x}_i$; $i=1, 2,...,k$ powinien charakteryzować się losowością. Stąd liczba serii w ciągu uzyskanym w wyniku porównania wartości $\bar{x}_i$ z ich medianą (Me_x) powinna być zbliżona do liczby serii charakterystycznej dla ciągu niezależnych obserwacji zmiennej losowej. Sprawdzianem w omawianym teście jest liczba serii uzyskanych w wyniku porównywania $\bar{x}_i$ z Me_x. Przy budowie obszaru krytycznego wykorzystywane są następujące równości:

$$P\{l \leqslant l_1\} = \tfrac{1}{2}\alpha \text{ i } P\{l < l_2\} = 1 - \tfrac{1}{2}\alpha,$$

gdzie:

l oznacza liczbę serii uzyskanych w danym ciągu, l_1, l_2 oznaczają wartości krytyczne odczytane z tablic rozkładu liczby serii dla poziomu istotności α (zob. [7]).

W świetle powyższych uwag widać, że gdy wyznaczona wartość l spełnia jedną z nierówności $l < l_1$ lub $l > l_2$, to hipotezę o stacjonarności rozważanego procesu należy odrzucić. W przypadku gdy $l_1 \leqslant l \leqslant l_2$, nie mamy podstaw do odrzucenia H_0.

Badanie stacjonarności procesu ze względu na wariancję wymaga analogicznego do przedstawionego wyżej postępowania z ciągiem wariancji s^2_i; $i=1, 2,...,k$. Proces uznajemy za stacjonarny, gdy nie mamy podstaw do odrzucenia hipotezy H_0 o jego stacjonarności zarówno ze względu na średnią, jak i na wariancję.

OCENA EFEKTYWNOŚCI KLASYCZNYCH METOD BADANIA STACJONARNOŚCI JEDNOWYMIAROWYCH CIĄGÓW LOSOWYCH

W tej części pracy oceniona zostanie praktyczna użyteczność medianowego testu serii i metody Boxa-Jenkinsa do badania stacjonarności procesów losowych. Statystyczna ocena wymienionych metod zostanie dokonana w oparciu o wyniki badań symulacyjnych.

Ocena obydwu sposobów postępowania wymagała generowania realizacji wybranych słabo stacjonarnych i niestacjonarnych procesów AR(p). Realizacje te otrzymywano w wyniku zastosowania następującej formuły:

$$x_t = 0, \qquad t=1, 2,..., p$$
$$x_t = \sum_{i=1}^{p} \alpha_{pi} x_{t-i} + \varepsilon_t \qquad t=p+1,...,n \qquad (14)$$

gdzie:

α_{pi}; i=1, 2,...,p — parametry autoregresji,

$\{\varepsilon_t;\ t=p+1,...,n\}$ — liczby pseudolosowe o rozkładzie normalnym N(0, 1).

Liczby pseudolosowe o rozkładzie normalnym N (0,1) uzyskiwano w wyniku zastosowania generatora wykorzystującego centralne twierdzenie graniczne. Opis tego generatora można znaleźć np. w [9, s. 84]. W przypadku, gdy ocenie poddawano metodę Boxa-Jenkinsa generowano realizacje następujących procesów:

AR(1): $\alpha_{11}=0,3$; $\alpha_{11}=0,4$; $\alpha_{11}=0,9$; $\alpha_{11}=1,0$; $\alpha_{11}=1,05$

AR(2): $\alpha_{21}=1$; $\alpha_{22}=-0,9$ — proces stacjonarny

$\alpha_{21}=1,1$; $\alpha_{22}=-0,1$ — proces niestacjonarny

AR(3): $\alpha_{31}=0,8$; $\alpha_{32}=0,51$; $\alpha_{33}=-0,378$ — proces stacjonarny

$\alpha_{31}=0,9$; $\alpha_{32}=0,52$; $\alpha_{33}=-0,42$ — proces niestacjonarny

Pojedynczy eksperyment polegał na 100-krotnym generowaniu realizacji jednego z wyżej wymienionych procesów. Na podstawie każdej realizacji dokonywano oceny wartości funkcji autokorelacji rzędu 1, 2,...,20 i okleślano rząd począwszy, od którego wartości funkcji autokorelacji okazały się statystycznie nieistotne. Rząd ten określany był na podstawie wyrażenia (9). W oparciu o każde 100 wyników wyznaczano średni rząd począwszy od którego wartości funkcji autokorelacji okazywały się statystycznie nieistotne.

Średni rząd obliczono na podstawie następującego wzoru:

$$sr=100^{-1}\sum_{100}^{i=1} li, \tag{15}$$

gdzie: l_i oznacza rząd począwszy od którego wartości funkcji autokorelacji okazały się nieistotne statystycznie w i-tym powtórzeniu.

Odchylenie standardowe wyników uzyskanych w danym eksperymencie obliczano według wzoru o postaci

$$sd=\sqrt{100^{-1}\sum_{i=1}^{100}(l_i-sr)^2}. \tag{16}$$

Dla kadżego procesu generowano realizacje o liczbie obserwacji n=50, 100, 200 i 500.

Wyniki badań przeprowadzonych w omówiony wyżej sposób zawarto w tabelach 1 i 2.

Analiza uzyskanych wyników pozwala zauważyć, że o ile w przypad-

Tab. 1. Ocena metody Boxa-Jenkinsa

The estimation of Box-Jenkins' method

Nazwa procesu	Liczba obserwacji							
	50		100		200		500	
	średni rząd	odchylenie standardowe	średni rząd	odchylenie standardowe	średni rząd	odchylenie standardowe	średni rząd	odchylenie standardowe
AR (1) $a_{11}=0,3$	3,63	2,76	5,53	5,61	6,08	5,68	8,36	6,59
AR (1) $a_{11}=0,4$	3,42	2,92	4,21	4,01	6,26	5,64	6,36	5,53
AR (1) $a_{11}=0,5$	3,39	2,82	4,42	4,02	5,37	4,60	5,46	4,99
AR (1) $a_{11}=0,6$	3,23	4,15	5,19	4,15	5,29	4,08	7,12	5,00
AR (1) $a_{11}=0,7$	3,59	2,94	5,19	3,66	5,73	3,46	7,46	3,53
AR (1) $a_{11}=0,8$	3,29	1,29	4,88	2,12	6,33	3,21	8,38	3,30
AR (1) $a_{11}=0,9$	3,69	1,62	6,13	2,42	8,96	2,52	12,52	3,01
AR (1) $a_{11}=1$	4,31	0,64	8,03	1,33	14,61	3,42	—[c]	—[c]
AR (1) $a_{11}=1,05$	4,79	0,52	8,01	0,36	—[c]	—[c]	—[c]	—[c]

Źródło: Obliczenia własne.

[c] — wszystkie oceny okazały się statystycznie istotne.

ku prób bardzo dużych (500 obserwacji) istnieje możliwość jednoznacznego stwierdzenia czy dany proces można uznać za stacjonarny, czy nie[4], to w przypadku prób o liczbie obserwacji n=50 i n=100 wyciągnięcie takiego wniosku jest praktycznie niemożliwe. Powyższa uwaga wynika z tego, że dla prób o liczbie obserwacji 50 i 100, średni rząd powyżej którego wartości funkcji autokorelacji są statystycznie nieistotne jest zbliżony zarówno dla procesów stacjonarnych, jak i niestacjonarnych. Jest on

[4] W przypadku, gdy generowano bardzo długie szeregi czasowe stanowiące realizacje procesów niestacjonarnych, to wszystkie wartości funkcji autokorelacji okazywały się statystycznie istotne.

wprawdzie dla procesów niestacjonarnych wyższy lecz tak nieznacznie, iż trudno jest wyciągnąć tutaj wnioski, które mogłyby być przydatne w praktyce badania stacjonarności. Uzyskane wyniki uświadamiają w pełni, jak bardzo zawodne jest wykorzystywanie metody Boxa i Jenkinsa do badania stacjonarności.

Kolejno przystąpiono do oceny efektywności badania stacjonarności przy wykorzystaniu medianowego testu serii. Celem badań było określenie mocy medianowego testu serii[5] w sytuacji, gdy stosowany jest on do badania stacjonarności procesu losowego. W tym przypadku zgodnie z zasadami podanymi na wstępie niniejszej części pracy generowano realizacje niestacjonarnych procesów losowych AR(1) z $\alpha_{11}=1$ i $\alpha_{11}=1,1$. Pojedynczy eksperyment polegał na 100 krotnym wygenerowaniu realizacji

Tab. 2. Ocena metody Boxa-Jenkinsa
The estimation of Box-Jenkins' method

Nazwa procesu	Liczba obserwacji							
	50		100		200		500	
	średni rząd	odchylenie standardowe	średni rząd	odchylenie standardowe	średni rząd	odchylenie standardowe	średni rząd	odchylenie standardowe
AR (2) $a_{21}=1$ $a_{22}=-0,09$	3,86	1,78	6,33	2,75	8,48	2,34	12,68	2,74
AR (2) $a_{21}=1,1$ $a_{22}=-0,1$	4,29	0,97	7,69	1,59	14,5	2,90	—[c]	—
AR (3) $a_{31}=0,8$ $a_{32}=0,51$ $a_{33}=-0,378$	3,93	1,00	6,31	1,50	9,39	2,74	11,8	3,72
AR (3) $a_{31}=0,9$ $a_{32}=0,52$ $a_{33}=-0,42$	4,43	0,75	7,99	1,30	15,18	3,21	—[c]	—

[c] — wszystkie oceny okazały się statystycznie istotne.
Źródło: obliczenia własne.

[5] Pod pojęciem mocy testu rozumie się tu prawdopodobieństwo odrzucenia hipotezy głoszącej stacjonarność procesu losowego w przypadku, gdy realizacja reprezentowała proces niestacjonarny.

Tab. 3. Moc testu stacjonarności
The power of the test of stationarity

Nazwa procesu		Moc testu	Nazwa procesu		Moc testu
AR (1) $a_{11}=1$	N=50 M=5	0,874	AR (1) $a_{11}=1,1$	N=50 M=5	0,856
	N=50 M=10	0,78			
	N=100 M=5	0,972		N=100 M=5	0,986
	N=100 M=8	0,858		N=100 M=8	0,846
	N=100 M=10	0,850		N=100 M=10	0,844
	N=150 M=5	1		N=150 M=10	0,99
	N=150 M=10	1		N=150 M=15	0,854
	N=150 M=15	0,892			
	N=200 M=20	0,872			

Źródło: obliczenia własne.

wymienionych wyżej procesów. Badanie stacjonarności przeprowadzano dla szeregów czasowych o różnej długości (n=50, 100, 150, 200) i przy założeniu różnej długości odcinków, na które szeregi te były dzielone. Wyniki badań przedstawiono w tabeli 3.

Rezultaty, które uzyskano w wyniku badania mocy testu stacjonarności w przypadku realizacji niestacjonarnych procesów AR(1) wydawały się być obiecujące. We wszystkich bowiem przypadkach prawdopodobieństwo odrzucenia fałszywej hipotezy zerowej było wysokie (0,85 — 1). Ponieważ przy postępowaniu polegającym na dwukrotnym zastosowaniu testu serii nie ma możliwości określenia prawdopodobieństwa popełnienia błędu I rodzaju, zdecydowano się na podjęcie badań symulacyjnych, które umożliwiłyby oszacowanie tego prawdopodobieństwa. Badania ograniczono jedynie do realizacji procesu AR(1) z $a_{11}=0,8$, gdyż już pierwsze wyniki przekreśliły praktyczną przydatność analizowanego testu. Okazało się bowiem, że dla n=50 i m=5 stacjonarność została stwierdzona jedynie w 21 na 100 powtórzeń, dla n=100, m=10 w 22 powtórzeniach, dla

Tab. 4. Empiryczny rozkład estymatora parametru α_{11}
Empirical distribution of the estimator of parameter α_{11}

Przedział klasowy	AR(1) $\alpha_{11}=-0,99$				AR(1) $\alpha_{11}=-1$				AR(1) $\alpha_{11}=-1,01$			
	n=50	n=100	n=200	n=300	n=50	n=100	n=200	n=300	n=50	n=100	n=200	n=300
poniżej (−1,175)	0[a]	0	0	0	0	0	0	0	0	0	0	0
(−1,175)—(−1,15)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,15)—(−1,125)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,125)—(−1,1)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,1)—(−1,075)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,075)—(−1,05)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,05)—(−1,025)	3	0	0	0	2	0	0	0	3	0	0	0
(−1,025)—(−1,00)	24	10	4	1	35	34	37	27	74	80	93	93
(−1,0)—(−0,975)	30	56	71	82	38	53	56	73	13	11	6	6
(−0,975)—(−0,95)	24	20	18	16	11	6	7	0	6	5	1	1
(−0,95)—(−0,925)	9	13	6	1	8	2	0	0	2	1	0	0
(−0,925)—(−0,9)	4	1	1	0	2	4	0	0	0	3	0	0
(−0,9)—(−0,875)	1	0	0	0	1	1	0	0	1	0	0	0
(−0,875)—(−0,85)	1	0	0	0	1	0	0	0	0	0	0	0
(−0,85)—(−0,825)	0	0	0	0	1	0	0	0	0	0	0	0
(−0,825)—(−0,8)	2	0	0	0	1	0	0	0	1	0	0	0
powyżej (−0,8)	2	0	0	0	0	0	0	0	0	0	0	0

a liczba ta określa liczbę powtórzeń, w których estymator parametru α_{11} przyjął wartość poniżej (−1,175).

Źródło: Obliczenia własne.

n=100, m=15 w 12 powtórzeniach, zaś dla n=300, m=15 zaledwie w 1 powtórzeniu. W świetle powyższych uwag możemy stwierdzić, że stosowane dotychczas metody badania stacjonarności procesów losowych są wysoce zawodne. O ile jednak metoda Boxa-Jenkinsa ma walory praktyczne w przypadku szeregów bardzo długich, to omówiony w pracy test stacjonarności jest pozbawiony takich walorów.

PROPOZYCJA METODY BADANIA STACJONARNOŚCI PROCESÓW LOSOWYCH

Podstawę proponowanej w nijnieszej pracy metody badania stacjonarności stanowi hipoteza, że w przypadku rozważania niestacjonarnego procesu losowego jako procesu AR(1), uzyskana klasyczną metodą najmniejszych kwadratów ocena parametru α_{11} przyjmuje wartość większą lub równą 1. Powyższą hipotezę weryfikowano w oparciu o wyniki badań

Tab. 5. Empiryczny rozkład estymatora parametru α_{11}

Empirical distribution of the estimator of parameter α_{11}

Przedział klasowy	AR (1) α_{11}=0,99				AR (1) α_{11}=1				AR (1) α_{11}=1,01			
	n=50	n=100	n=200	n=300	n=50	n=100	n=200	n=300	n=50	n=100	n=200	n=300
poniżej 0,8	12	0	0	0	7	1	0	0	9	0	0	0
0,8—0,825	7	0	1	0	4	1	0	0	4	0	0	0
0,825—0,85	6	2	0	0	10	4	0	0	2	2	0	0
0,85—0,875	7	5	0	0	12	4	1	0	3	1	0	0
0,875—0,9	11	3	3	1	8	4	1	0	8	0	1	0
0,9—0,925	6	9	2	0	14	12	1	0	15	1	0	0
0,925—0,95	22	27	12	14	20	20	6	6	9	8	0	0
0,95—0,975	16	28	37	34	14	23	31	21	15	15	5	3
0,975—1	10	24	44	49	8	27	55	64	17	31	10	7
1—1,025	3	2	1	2	3	4	5	9	15	42	84	90
1,025—1,05	0	0	0	0	0	0	0	0	2	0	0	0
1,05—1,075	0	0	0	0	0	0	0	0	1	0	0	0
1,075—1,1	0	0	0	0	0	0	0	0	0	0	0	0
1,1—1,125	0	0	0	0	0	0	0	0	0	0	0	0
1,125—1,15	0	0	0	0	0	0	0	0	0	0	0	0
1,15—1,175	0	0	0	0	0	0	0	0	0	0	0	0

Źródło: Obliczenia własne.

symulacyjnych przeprowadzonych dla wybranych stacjonarnych i niestacjonarnych autoregresyjnych procesów losowych.

Procesy te dobierano w ten sposób, by pierwiastki odpowiadających im równań charakterystycznych znajdowały się w pobliżu lub na granicy stacjonarności. Zgodnie z formułą (14) generowano realizacje następujących procesów losowych:

AR(1): $\alpha_{11}=0,99$; $\alpha_{11}=1$; $\alpha_{11}=1,01$

$\alpha_{11}=-0,99$; $\alpha_{11}=-1$; $\alpha_{11}=-1,01$

AR(2): $\alpha_{21}=1,1$; $\alpha_{22}=-0,1$ — proces niestacjonarny (pierwiastek na granicy stacjonarności),

$\alpha_{21}=1,1$; $\alpha_{22}=-0,09$ — proces niestacjonarny,

$\alpha_{21}=1,1$; $\alpha_{22}=-0,11$ — proces stacjonarny,

$\alpha_{21}=-0,21$; $\alpha_{22}=0,808$ — proces niestacjonarny,

α_{21}: $-0,2$; $\alpha_{22}=0,8$ — proces niestacjonarny (pierwiastek na granicy stacjonarności

$\alpha_{21}=-0,19$; $\alpha_{22}=0,792$ — proces stacjonarny

AR(3): $\alpha_{31}=1,1$; $\alpha_{32}=0,32$; $\alpha_{33}=-0,42$ — proces niestacjonarny (pierwiastek na granicy stacjonarności)

$\alpha_{31}=1,11$; $\alpha_{32}=0,319$; $\alpha_{33}=-0,4242$ — proces niestacjonarny

Tab. 6. Empiryczny rozkład estymatora parametru a_{11}
Empirical distribution of the estimator of parameter a_{11}

Przedział klasowy	AR (2) $a_{21}=1,1$ $a_{22}=-0,11$				AR (2) $a_{21}=1,1$ $a_{22}=-0,1$				AR (1) $a_{21}=1,1$ $a_{22}=-0,09$			
	n=50	n=100	n=200	n=300	n=50	n=100	n=200	n=300	n=50	n=100	n=200	n=300
poniżej 0,8	11	0	0	0	4	0	0	0	3	0	0	0
0,8—0,825	2	1	0	0	4	1	0	0	1	0	0	0
0,825—0,85	8	0	0	0	4	1	0	0	0	0	0	0
0,85—0,875	4	4	0	0	2	3	0	0	8	0	0	0
0,875—0,9	14	4	0	0	13	3	0	0	6	2	0	0
0,9—0,925	14	13	3	1	20	9	2	0	8	4	0	0
0,925—0,95	18	25	5	3	15	11	6	0	11	1	1	0
0,95—0,975	15	29	30	28	16	33	27	20	17	9	4	0
0,975—1,0	9	22	61	68	15	37	57	79	23	22	7	4
1,0—1,025	5	2	1	0	7	2	8	1	20	61	88	96
1,025—1,05	0	0	0	0	0	0	0	0	3	1	0	0
1,05—1,075	0	0	0	0	0	0	0	0	0	0	0	0
1,075—1,1	0	0	0	0	0	0	0	0	0	0	0	0
1,1—1,125	0	0	0	0	0	0	0	0	0	0	0	0
1,125—1,15	0	0	0	0	0	0	0	0	0	0	0	0
1,15—1,175	0	0	0	0	0	0	0	0	0	0	0	0

Źródło: Obliczenia własne.

$\alpha_{31}=1{,}11$; $\alpha_{32}=0{,}319$; $a_{33}=-0{,}4242$ — proces niestacjonarny
$\alpha_{31}=1$; $\alpha_{32}=0{,}33$; $a_{33}=-0{,}37$ — proces stacjonarny

Pojedynczy eksperyment polegał na 100-krotnym wygenerowaniu realizacji jednego z wymienionych wyżej procesów. Na podstawie każdej realizacji w oparciu o metodę najmniejszych kwadratów szacowano parametr α_{11} modelu AR(1). Uwzględniając wyniki dla wszystkich 100 powtórzeń uzyskiwano empiryczne rozkłady estymatora parametru α_{11}. Badania przeprowadzono dla prób o liczebnościach n=50, 100, 200 i 300. Uzyskiwane wyniki zamieszczono w tabelach 4, 5, 6, 7 i 8.

Szczegółowa analiza informacji zawartych w tabelach 4—8 pozwala stwierdzić, że dla 200 i 300 elementowych realizacji procesów niestacjonarnych o pierwiastkach równania charakterystycznego znajdujących się wewnątrz okręgu o promieniu jednostkowym w około 90% eksperymentów uzyskiwano ocenę α_{11} powyżej wartości 1 lub poniężj wartości — 1. Dla prób 50 i 100 elementowych wyniki te nie były już tak korzystne i kształtowały się odpowiednio na poziomie około 15—20% oraz 42—60%. Widać jednak wyraźnie, że w przypadku, gdy bezwzględna wartość z oceny parametru α_{11} jest większa lub równa 1, to proces reprezentowany przez daną próbę można zakwalifikować do klasy niestacjonarnych. Rozkład estymatora

Tab. 7. Empiryczny rozkład estymatora parametru a_{11}
Empirical distribution of the estimator of parameter a_{11}

Przedział klasowy	AR (2) $a_{21}=-0,21$ $a_{22}=0,808$				AR (2) $a_{21}=-0,2$ $a_{22}=0,8$				AR (2) $a_{21}=-0,19$ $a_{22}=0,792$			
	n=50	n=100	n=200	n=300	n=50	n=100	n=200	n=300	n=50	n=100	n=200	n=300
poniżej (−1,175)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,175)—(−1,15)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,15)—(−1,125)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,125)—(−1,1)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,1)—(−1,075)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,075)—(−1,05)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,05)—(−1,025)	0	0	0	0	0	0	0	0	0	0	0	0
(−1,025)—(−1,0)	35	61	75	90	4	7	2	3	1	0	0	0
(−1,0)—(−0,975)	36	16	10	5	23	36	40	48	6	10	2	1
(−0,975)—(−0,95)	7	7	2	2	19	13	21	20	7	8	10	15
(−0,95)—(−0,925)	7	3	4	2	9	7	14	7	15	15	19	16
(−0,925)—(−0,9)	2	1	3	0	4	8	7	5	10	9	13	19
(−0,9)—(−0,875)	2	3	3	0	2	7	2	5	9	10	10	19
(−0,875)—(−0,85)	1	1	0	0	6	5	1	4	10	8	8	8
(−0,85)—(−0,825)	2	2	1	1	3	2	4	3	3	7	9	7
(−0,825)—(−0,8)	1	3	0	0	4	2	2	1	2	4	4	4
powyżej (−0,8)	7	3	2	0	26	13	0	4	37	29	25	11

Źródło: Obliczenia własne.

uzyskany dla procesów stacjonarnych wyraźnie wskazuje, że przy wykorzystaniu powyższej uwagi do badania stacjonarności prawdopodobieństwo popełnienia błędu polegającego na zakwalifikowaniu procesu stacjonarnego do klasy niestacjonarnych jest bardzo małe i należy do przedziału od 0 do 0,05. Łatwo zauważyć, że w przypadku procesów niestacjonarnych dla prób o liczbie obserwacji 200 i 300 prawdopodobieństwo tego, że bezwzględna wartość oceny parametru α_{11} jest mniejsza od 0,95 jest małe i należy do przedziału od 0 do 0,1. Dla 100 prób elementowych małe jest prawdopodobieństwo tego, że bezwzględna wartość oceny parametru α_{11} jest mniejsza od 0,925. Należy ono do przedziału od 0 do 0,2. W przypadku prób 50 elementowych trudno jest mówić o jakiejkolwiek prawidłowości.

W świetle powyższych uwag można stwierdzić, że zarówno dla prób małych, jak i dużych wystąpienie większej lub równej 1, bądź mniejszej lub równej −1 wartości estymatora parametru a_{11} pozwala stwierdzić niestacjonarność procesu reprezentowanego przez dany szereg czasowy z niewielkim prawdopodobieństwem popełnienia błędu. W przypadku prób o liczbie obserwacji większej lub równej 100 wystąpienie bezwzglę-

Tab. 8. Empiryczny rozkład estymatora parametru a_{11}
Empirical distribution of the estimator of parameter a_{11}

Wartość estymatora	AR (3) $a_{31}=1$, $a_{32}=0,33$, $a_{33}=-0,378$				AR (3) $a_{31}=1,1$, $a_{32}=0,32$, $a_{33}=-0,42$				AR (3) $a_{31}=1,11$, $a_{32}=0,32$, $a_{33}=-0,42$			
	n=50	n=100	n=200	n=300	n=50	n=100	n=200	n=300	n=50	n=100	n=200	n=300
poniżej 0,8	2	0	0	0	0	0	0	0	0	0	0	0
0,8—0,825	1	0	0	0	2	0	0	0	0	0	0	0
0,825—0,85	1	0	0	0	1	0	0	0	3	0	0	0
0,85—0,875	8	3	0	0	4	0	0	0	5	0	0	0
0,875—0,9	11	3	0	0	6	0	0	0	2	0	0	0
0,9—0,925	16	17	2	0	8	2	0	0	6	0	0	0
0,925—0,95	19	21	21	13	21	6	0	0	12	3	0	0
0,95—0,975	23	41	67	79	19	24	8	0	28	6	1	0
0,975—1	14	15	10	8	21	48	66	78	24	32	7	4
1—1,025	3	0	0	0	13	19	26	22	11	52	92	96
1,025—1,05	1	0	0	0	4	1	0	0	8	7	0	0
1,05—1,075	1	0	0	0	1	0	0	0	1	0	0	0
1,075—1,1	0	0	0	0	0	0	0	0	0	0	0	0
1,1—1,125	0	0	0	0	0	0	0	0	0	0	0	0
1,125—1,15	0	0	0	0	0	0	0	0	0	0	0	0
1,15—1,175	0	0	0	0	0	0	0	0	0	0	0	0

Źródło: Obliczenia własne.

dnej wartości estymatora parametru a_{11} na poziomie niższym od 0,925 pozwala zakwalifikować proces reprezentowany przez dany szereg czasowy do klasy stacjonarnych. Przedstawione powyżej wnioski stanowią przekonywającą podstawę proponowanej w pracy metody badania stacjonarności.

Metoda ta nie rozstrzyga problemu stacjonarności bądź niestacjonarności w sytuacji, gdy bezwzględna wartość oceny parametru a_{11} należy do przedziału (0,925; 1). Proponuje się, by w takiej sytuacji proces był niestacjonarny, czyli, aby badaniu stacjonarności zgodnie z przedstawionymi wyżej uwagami poddany został szereg pierwszych bądź kolejnych różnic.

ZAKOŃCZENIE

Przeprowadzone w pracy badania upoważniają do sformułowania następujących wniosków:

1. Medianowy test serii nie daje możliwości wiarygodnego rozstrzygania problemu stacjonarności bądź niestacjonarności procesów losowych;

2. Rozróżnienie realizacji procesów losowych stacjonarnych i niestacjonarnych w oparciu o analizę zachowania wartości estymatora funkcji autokorelacji jest w przypadku realizacji 50, 100 i często nawet 200 elementowych praktycznie niemożliwe;

3. W przypadku, gdy uzyskane w oparciu o metodę najmniejszych kwadratów oceny parametru a_{11} procesu AR(1) charakteryzują się bezwzględną wartością większą lub równą 1, to proces należy zaliczyć do klasy niestacjonarnych;

4. W przypadku, gdy bezwzględna wartość oceny parametru a_{11} jest mniejsza od 0,925, to proces można zaliczyć do klasy stacjonarnych;

5. Jeżeli bezwzględna wartość oceny parametru a_{11} należy do przedziału (0,925; 1) nie ma możliwości określenia, do jakiej klasy należy proces. W takim przypadku należy proces ten rozważać jako proces niestacjonarny.

LITERATURA

1. Anderson O. D.: A New Approach to ARMA Modelling: Some Comments. Analysing Time Series, North-Holland Publishing Company—Amsterdam, 1980.
2. Anderson T. W.: Statisticzeskij analiz wriemiennych rjadow. Mir, Maskwa 1976.
3. Beamish N., Priestley H. B.: A Study of Autoregressive and Window Spectral Estimation. Applied Statistics, 1981, 30, nr 1, s. 41—58.
4. Bendat J. S., Piersol A. G.: Metody analizy i pomiaru sygnałów losowych, PWN, Warszawa 1976.
5. Bora-Senta E., Kounias S.: Parameter Estimation and Order Determination of Autoregressive Models, Analysing Time Series, North-Holland Publishing Company, 1980.
6. Box G. E. P., Jenkins G. M.: Analiza szeregów czasowych, PWN, Warszawa 1983.
7. Domański Cz: Statystyczne testy nieparametryczne, PWE, Warszawa 1979.
8. Fishman G. S.: Symulacja komputerowa, pojęcia i metody. PWE, Warszawa 1981.
9. Zieliński R.: Generatory liczb losowych, WNT, Warszawa 1979.

РЕЗЮМЕ

В статье обсуждаются применяемые в настоящее время методы исследования стационарности одномерных случайных процессов, т. е. метод Бокса-Дженкинса и медианный тест серии. Опираясь на результаты симуляционных исследований, автор обнаруживает несовершенство названных методов. В статье предлагается процедура, которая представляется конкурентной по отношению к названным выше. Предлагаемая процедура представляет собой результат симуляционных исследований, проведенных для широкого класса стационарных и нестационарных авторегрессивных процессов.

SUMMARY

The paper presents an evaluation of the methods which so far have been applied to study stationarity of one-dimensional Stochastic processes. These methods are Box-Jenkins' method and a median series test. The results of simulation studies prove the deceptiveness of the methods under discussion. A procedure was put forward which seems to be competivite in relation to the ones mentioned above. This procedure is a result of simulation studies conducted for a wide class of stationary and non-stationary auto-regressive processes.